DBM1: UEFA Champions League 2016 – 2022

IST – Database and Data Mining

1st Semester 2021/2022

14/11/2022

Written by:

Andrei-Marian Enache, andrei-marian.enache@insa-lyon.fr

Maria Minodora Mares, maria-minodora.mares@insa.lyon.fr

I.   Introduction

This document serves as a report about the work we have done to build the final project for the Database and Data Mining course. It includes a description of the chosen dataset, details about our design decisions and an overview of the implementation steps for both the database itself and the queries that were written in SQL and Relational Algebra.

II.   Information About The Dataset

The dataset we decided to work with contains data on the UEFA Champions League seasons between 2016 and 2022. It can be obtained through Kaggle – dataset, and it provides information about players, teams, managers, matches, stadiums, results and statistics.

III.   ER Diagram
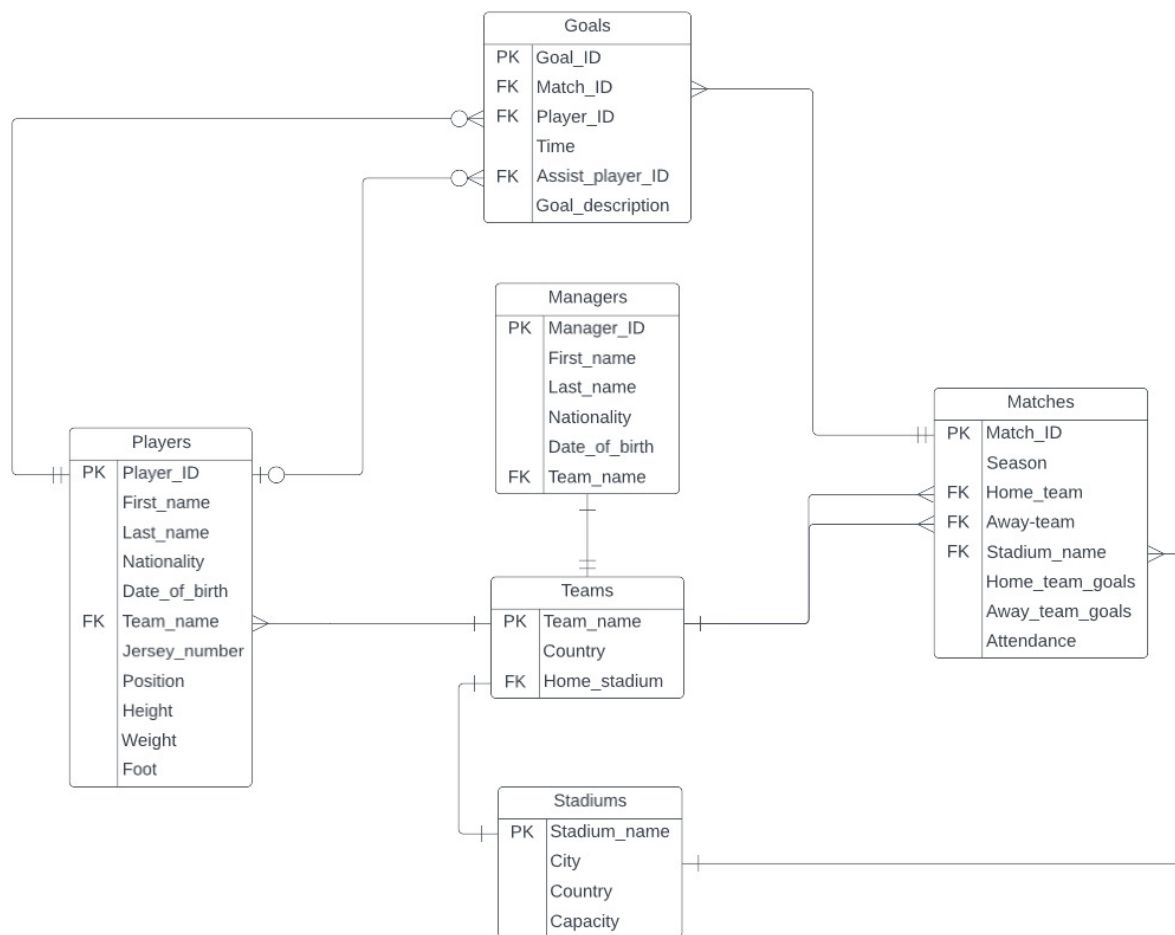
Throughout the design of our ER diagram, we approached things from small to big, so the first two entities to be created were the *players* and the *managers*. Since these two are the fundamental pieces of a team, we expanded our diagram by adding the *teams* entity next. The *matches* entity followed, being the hardest to model because we had to associate it two times with the *teams* – once for the home team and once for the away team. Each team has a home stadium and each

match takes place on a stadium, so the creation of a *stadiums* entity, connected with *teams* as well as with *matches*, seemed to be the next helpful step in designing our diagram. The last entity that was implemented was the *goals*, which revealed an association with *matches* as we need to know in which match the goal was scored, but also a double connection with the *players* to record the player who scored and the one who gave the assist.

IV.   Implementation of The Database

With the dataset at hand and the ER diagram built, we used PostgreSQL to kick off the actual implementation of the database.

First, we created all the entity tables with their respective attributes and primary keys (as designed in the ER diagram). From that, multiple foreign key constraints were set: a match has a home team, an away team and a stadium, a goal is associated with the player who scored, the player who assisted and a match, each player and each manager are related to a team, each team has their own stadium. Finally, we built the relationship between the entities. For associations inheriting a one to many relationship (Teams – Players, Teams – Matches, Matches – Goals, Players – Goals), we added a variable, the primary key of the one side, in the table of the many side. For one to one relationships (Teams – Managers, Teams – Stadiums), we chose the side to which we brought the foreign key depending on the situation that will make the resolution of our queries easier.

We had issues with the implementation of the foreign keys in PostgreSQL because, when importing the csv files for tables with foreign keys, an error arises. We didn't manage to fix that, so we implemented it as a normal attribute.

V.   Populating The Database

After building all the tables in PostgreSQL, we imported them as csv files(previously downloaded from Kaggle) using pgAdmin. The order was really important because we need to first populate the tables that do not rely on others and then the ones referencing information already inserted in the database. With that being said, our import order was the following: stadiums, teams, players, managers, matches and goals.

VI.   SQL Queries

These are the questions that we wanted to find an answer for within our database:

1) How many players are from Italy?
2) Display the descriptions of all the goals scored by Cristiano Ronaldo in 2017/2018 season.
3) Display the home team, away team and final score of matches in which the home team manager is born after 1975.
4) Display the details of the players and number of goals scored in 2020/2021 season.
5) What are the stadiums where more than 20 matches were played?

It was challenging to translate the queries into the SQL format. Nonetheless, all of them were achieved. The SQL implementation of the queries can be found in queries.sql.

VII.   Conclusion

We built a database through which we can analyze data on Champions League competitions from 2016 to 2022. It was easier to grasp all the concepts we studied throughout the lectures thanks to this hands-on experience which exposed us to using different tools (PostgreSQL) in order to answer specific questions.